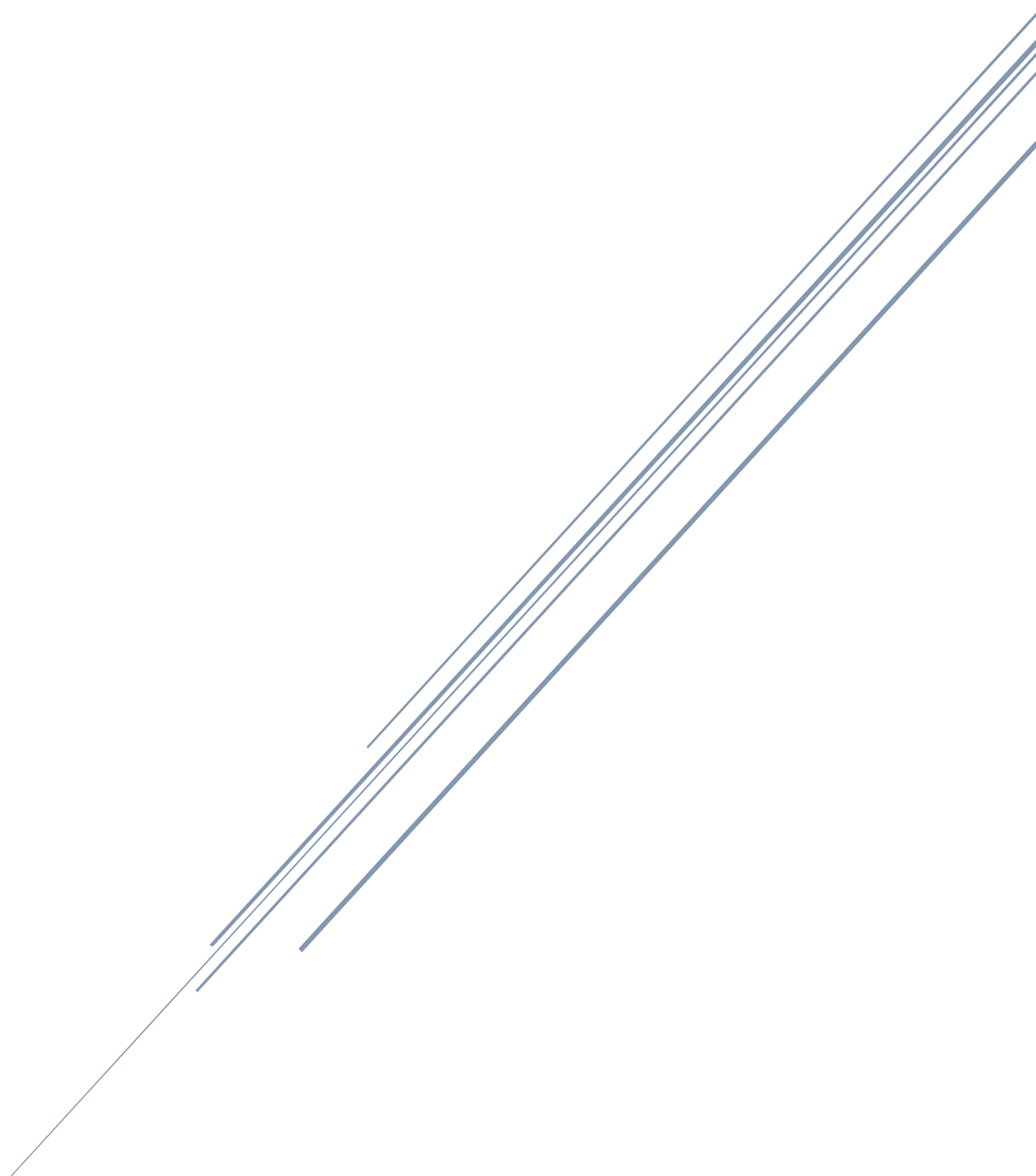


# HUSTOJ 数据分析与可视化工具

## 最终用户使用说明书



东南大学生物科学与医学工程学院 夏小俊课题组

祝云麓 童澄达 赵作翰

# 目录

目录 .....	1
一、软件说明.....	2
1.1 软件功能说明 .....	2
1.2 软件的开发、测试与运行环境.....	2
1.2.1 软件的开发环境 .....	2
1.2.2 软件的测试环境 .....	2
1.2.3 软件的运行环境 .....	3
1.3 开发团队信息 .....	3
1.4 开放源代码许可.....	4
1.5 说明书更新记录.....	4
二、软件的安装与卸载.....	5
2.1 软件的安装.....	5
2.1.1 预配置.....	5
2.1.2 安装运行环境.....	5
2.1.3 执行安装程序.....	6
2.2 软件的卸载.....	8
三、软件的使用 .....	9
3.1 软件的启动.....	9
3.2 软件的使用 .....	10
3.3 指定分析起讫时间 .....	11
3.4 更新数据.....	12
3.5 导出数据.....	13
3.5.1 导出原始数据.....	13
3.5.2 导出学生数据 .....	14
3.5.3 导出题目数据 .....	15
3.5.4 导出系统数据 .....	16
四、附录.....	17
4.1 软件原理.....	17

# 一、软件说明

## 1.1 软件功能说明

本应用程序系为使用使用《东南大学生物科学与医学工程学院在线评测（Online Judge, OJ）系统》进行主要面向大学一年级同学开设的必修课程《程序设计与算法语言（电类）》的教学的教师同志研发的工具软件。主要目的为方便教师对参与在线评测的学生的程序设计能力进行综合分析。该软件可对来自《东南大学生物科学与医学工程学院在线评测（Online Judge, OJ）系统》（网络地址为 <http://oj.bmeonline.cn/>）的学生提交和评测日志进行抓取和分析，并从学生、题目两个维度向用户提供可供分析的数据。用户可基于软件返回的数据对题目难度、学生参与性、学生程序设计能力等进行评估和分析。软件同样可以基于对学生各项指标的聚类分析给出针对学生的学业成绩预警。

## 1.2 软件的开发、测试与运行环境

### 1.2.1 软件的开发环境

本应用程序在表 1.1 所述环境中被开发。

表 1.1 软件的开发环境

项目	值
操作系统	Microsoft Windows 10 企业版 2016 长期服务版 (10.0.14393.3181)
开发工具	Microsoft Visual Studio 2013 Ultimate
	Microsoft .NET Framework 4.5
运行环境	MathWorks MATLAB Runtime R2018b
	Microsoft Visual C++ 2012 Redistributable (x86/x64)
	Microsoft Visual C++ 2015 Redistributable (x86/x64)
额外安装	-

### 1.2.2 软件的测试环境

本应用程序在表 1.2 所述环境下测试通过。

表 1.2 软件的测试环境 1

项目	值
操作系统	Microsoft Windows 10 企业版 2016 长期服务版 (10.0.14393.3181)
	Microsoft .NET Framework 4.5
运行环境	MathWorks MATLAB Runtime R2018b
	Microsoft Visual C++ 2012 Redistributable (x86/x64)
	Microsoft Visual C++ 2015 Redistributable (x86/x64)
额外安装	-

本应用程序在表 1.3 所述环境下测试通过。

表 1.3 软件的测试环境 2

项目	值
操作系统	Microsoft Windows Embedded Standard 7 SP1 (6.1.7601)
	Microsoft .NET Framework 4.0
运行环境	MathWorks MATLAB Runtime R2018b
	Microsoft Visual C++ 2015 Redistributable (x86/x64)
额外安装	适用于 Microsoft .NET Framework 4.0 的修补程序 KB2468871

### 1.2.3 软件的运行环境

#### 1.2.3.1 软件环境

本软件可在以下操作系统版本上运行：

Microsoft Windows XP Service Pack 3 (x86/x64)  
Microsoft Windows Server 2003 Service Pack 2  
Microsoft Windows Vista Service Pack 2 (x86/x64)  
Microsoft Windows Server 2008 Service Pack 2  
Microsoft Windows Server 2008 R2 Service Pack 1  
Microsoft Windows 7 Service Pack 1 (x86/x64)  
Microsoft Windows 8 (x86/x64)  
Microsoft Windows 8.1 (x86/x64)  
Microsoft Windows Server 2012  
Microsoft Windows Server 2012 R2  
Microsoft Windows 10  
Microsoft Windows Server 2016  
Microsoft Windows Server 2019

本软件要求以下运行环境为必选：

Microsoft .NET Framework 4.0

本软件要求为运行环境安装以下修补程序：

适用于 Microsoft .NET Framework 4.0 的修补程序 KB2468871

本软件在安装了以下运行环境中的一者或全部后可获得最佳体验

MathWorks MATLAB Runtime R2018b （仅限 64 位操作系统）

MathWorks MATLAB R2018b （仅限 64 位操作系统）

#### 1.2.3.2 硬件环境

本软件暂未测定最低硬件需求。仅建议用户在具有 1280x1024 或更高分辨率的显示设备的计算机上使用本应用程序。

## 1.3 开发团队信息

本软件由东南大学夏小俊课题组开发。

主程序员：童澄达

算法设计与数据更新程序开发：祝云簏

## 1.4 开放源代码许可

本软件遵照 MIT License 许可协议, 在 <http://share.bmeonline.cn/admin-and-development/oj-p2s> 上开放软件源代码。

### 附文 1.1 MIT License 许可协议

MIT License

Copyright (c) 2019 东南大学夏小俊课题组

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

## 1.5 说明书更新记录

本软件使用说明书的更新记录如表 1.4 所示。

表 1.4 本软件使用说明书更新记录

日期	修订版本	更新人	更新说明
2019 年 09 月 22 日	fbl-release.190922-2110	童澄达	首次发行。
2020 年 01 月 28 日	fbl-release.200128-1445	童澄达	添加有关数据导出功能的说明。 小幅格式修订。

## 二、软件的安装与卸载

### 2.1 软件的安装

#### 2.1.1 预配置

开始安装前,请确保您的系统已将 [非 Unicode 程序目前使用的语言] 设为 [中文 (简体, 中国)]。您可以在 [控制面板] > [地区] > [系统管理] 中找到这一选项。

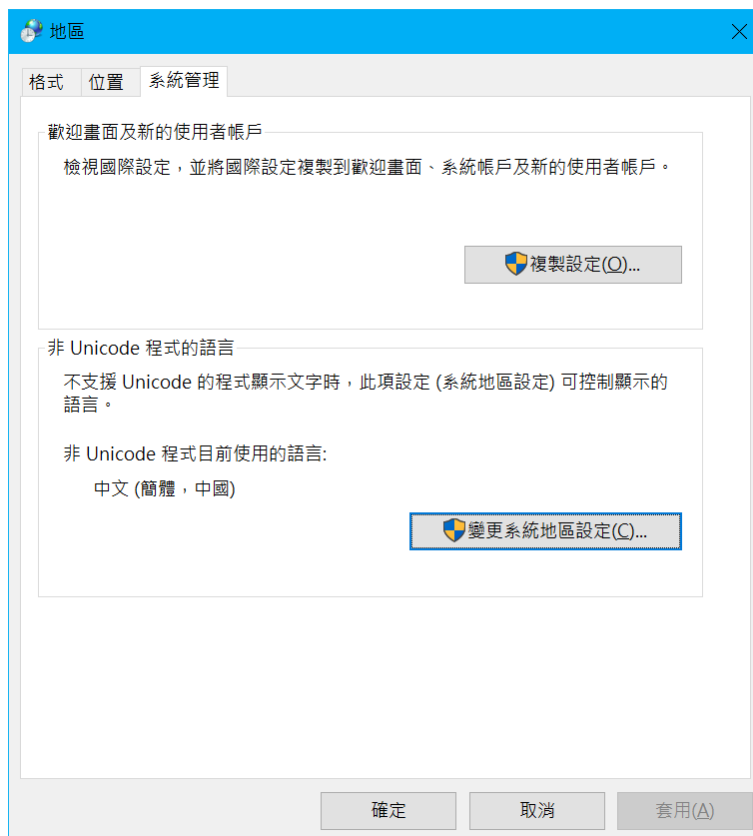


图 2.1 配置非 Unicode 程序使用的语言

#### 2.1.2 安装运行环境

尽管软件会尝试为您配置运行环境,但是如果您的系统中尚未安装 Microsoft .NET Framework 4.0, 仍然建议您手动安装软件所需的运行环境。

您可以从 Microsoft 下载中心下载适合您的操作系统版本的 Microsoft .NET Framework 4.0 及其修补程序 KB2468871 的安装程序文件。如果您期望获得完整体验, 请考虑在 64 位系统上安装此软件, 并从 MathWorks 网站下载 MathWorks MATLAB Runtime R2018b。

您可以从此处下载 Microsoft .NET Framework 4.0:

<https://www.microsoft.com/zh-cn/download/details.aspx?id=17718>

您可以从此处下载适用于 Microsoft .NET Framework 4.0 的修补程序 KB2468871:

<https://www.microsoft.com/en-us/download/details.aspx?id=3556>

若有需要, 您可以手动下载 Microsoft .NET Framework 4.5:

<https://www.microsoft.com/zh-CN/download/details.aspx?id=30653>

您可以从此处下载 MathWorks MATLAB Runtime R2018b:

<https://ww2.mathworks.cn/products/compiler/matlab-runtime.html>

### 2.1.3 执行安装程序

请运行分发给您的安装程序（它通常被打包为一个 EXE 文件），并根据指引完成安装。



图 2.2 安装程序的初始界面



图 2.3 配置快捷方式位置



图 2.4 安装



图 2.5 完成安装

完成安装后，您可以在 [桌面] 和/或 [开始]菜单中找到用于启动软件的快捷方式。



## 2.2 软件的卸载

您可以在 [控制面板] > [程序和功能] 或 [设定] > [系统] > [应用和功能]（对于更高版本的 Windows 10 操作系统，这个位置可能为[设定] > [应用] > [应用和功能]）中找到用于卸载软件的链接。

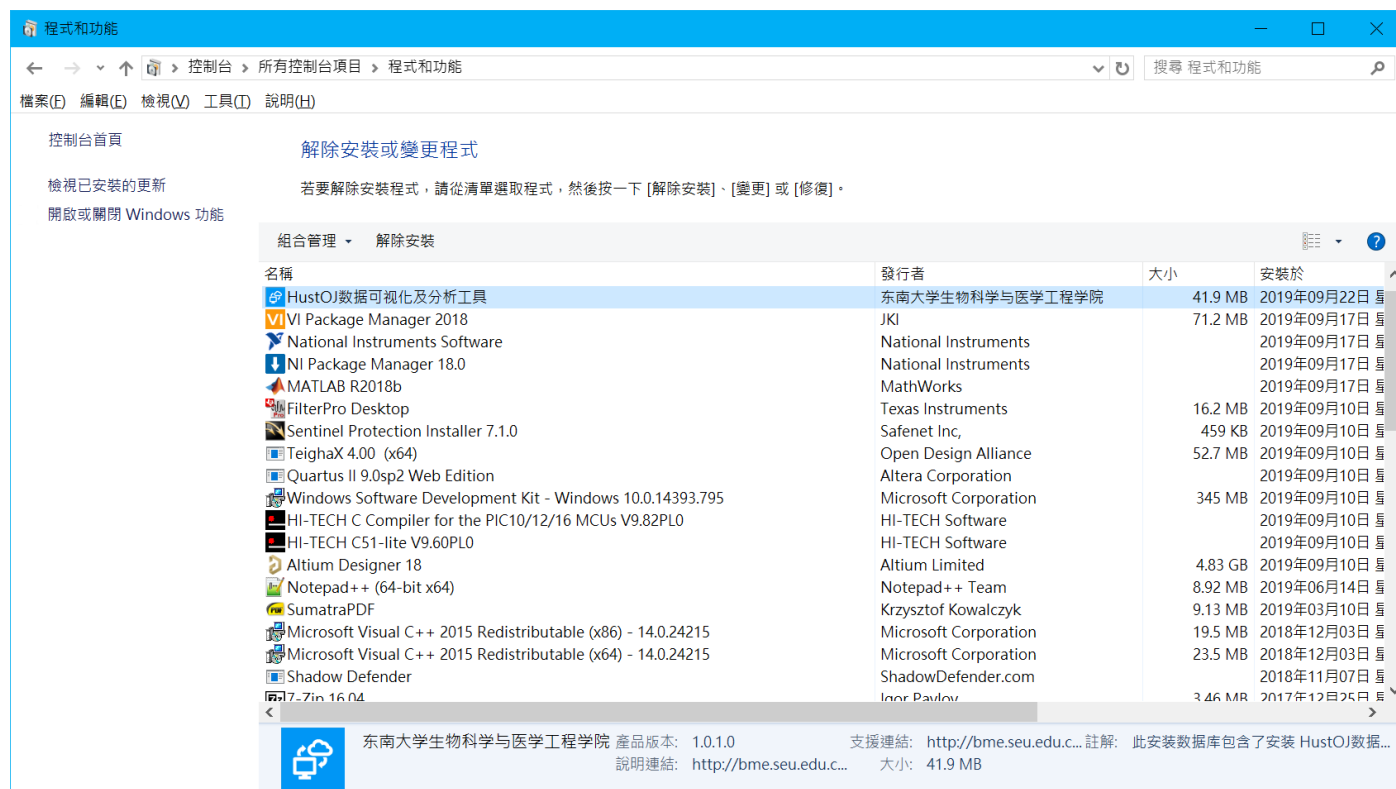


图 2.6 [控制面板] > [程序和功能] 中的卸载链接

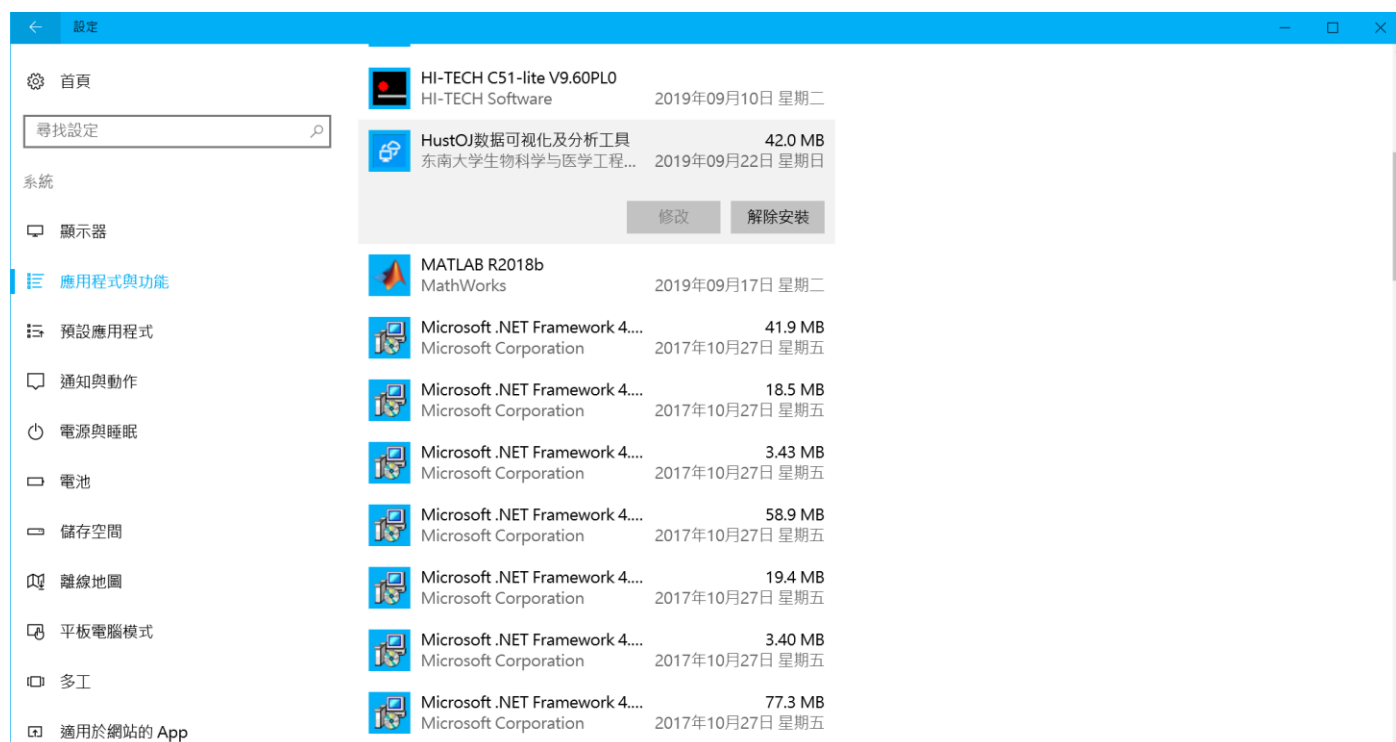


图 2.7 [设定] 应用中的卸载链接

# 三、软件的使用

## 3.1 软件的启动

软件启动后，会自动校验数据，并对数据进行分析。



图 3.1 软件启动后分析数据

在该过程中，软件会捕获发生的错误，并将其回报给用户。部分常见错误错误可能的诱发原因及其简单排查方案如表 3.1 所示。

表 3.1 常见问题及其排查指南

错误内容	可能的原因及解决方案
弹出错误窗口，显示.NET Assembly 错误	可能的原因：您安装了 Microsoft .NET Framework 4.0，但没有安装其修补程序 KB2468871。 可能的解决方案：请从 Microsoft 下载中心下载的修补程序 KB2468871。
弹出错误窗口，提示无法定位输入点 ucrtbase。	可能的原因：Microsoft 通用 C 运行库未被安装或发生错误。 可能的解决方案：请先通过 Windows Update 或其他渠道更新您的操作系统，并从 Microsoft 下载中心下载并安装 Visual C++ Redistributable for Visual Studio 2015。

(续表)

错误内容	可能的原因及解决方案
数据分析后，弹出错误窗口，提示 MATLAB Runtime 发生错误，且所有项目均显示为“出错或被跳过”。	可能的原因：MATLAB Runtime R2018b 未被安装或发生问题。 可能的解决方案：请安装或重新安装 MATLAB Runtime R2018b。
数据分析后，弹出错误窗口，提示 MATLAB Runtime 发生错误，但仅部分项目显示为“出错或被跳过”。	可能的原因：数据存在问题 可能的解决方案：请在进入程序后重新框定分析时间段。

3.2 软件的使用

本软件的主窗口分为两部分，左侧列出学生学号或题号，右侧提供各项数值。



图 3.2 软件主窗口

您可以在搜索框中进行过滤，该过滤不区分大小写。  
将鼠标指针移动到软件呈现的图表上可以显示当前指针所指向的图表分区的内容。  
默认的学生聚类结果为未分级，您需要点击该文本并在弹出的窗口中手动指定分级。

请注意：  
此处展现的软件版本为一般客户版，若分发给您的版本为终端机版或演示版，则软件运行后需要输入工程师凭据或维护凭据才能退出，且任务管理器等高级系统功能可能被屏蔽或无法正常运转，所有对系统的改动将在程序退出时被还原。若程序异常退出导致未撤销改动，请联系系统管理员。

### 3.3 指定分析起讫时间

您可以在软件顶部的工具栏自定义分析起讫时间。您需要按 [更新数据] 按钮才能更新数据。

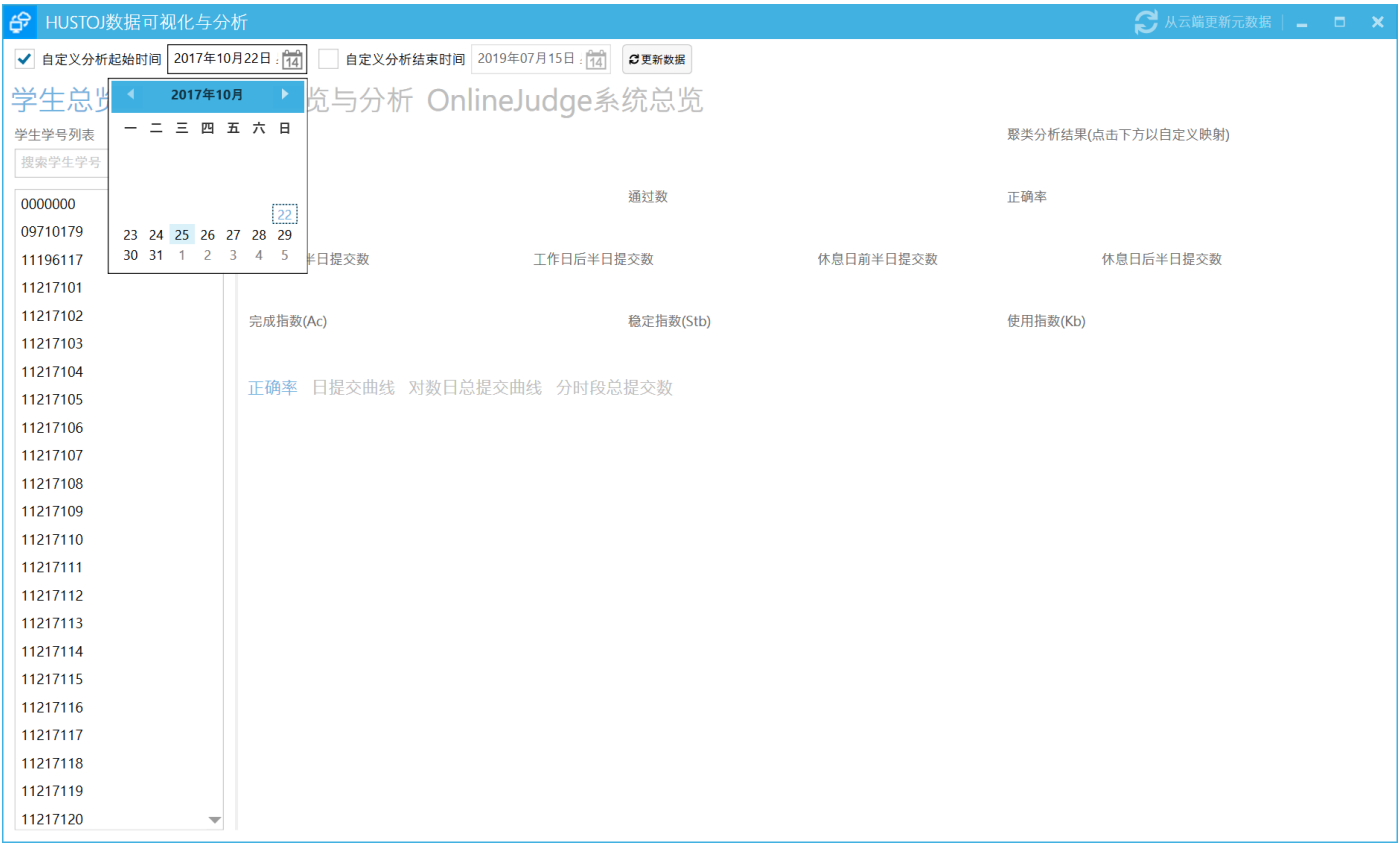


图 3.3 自定义分析起始时间

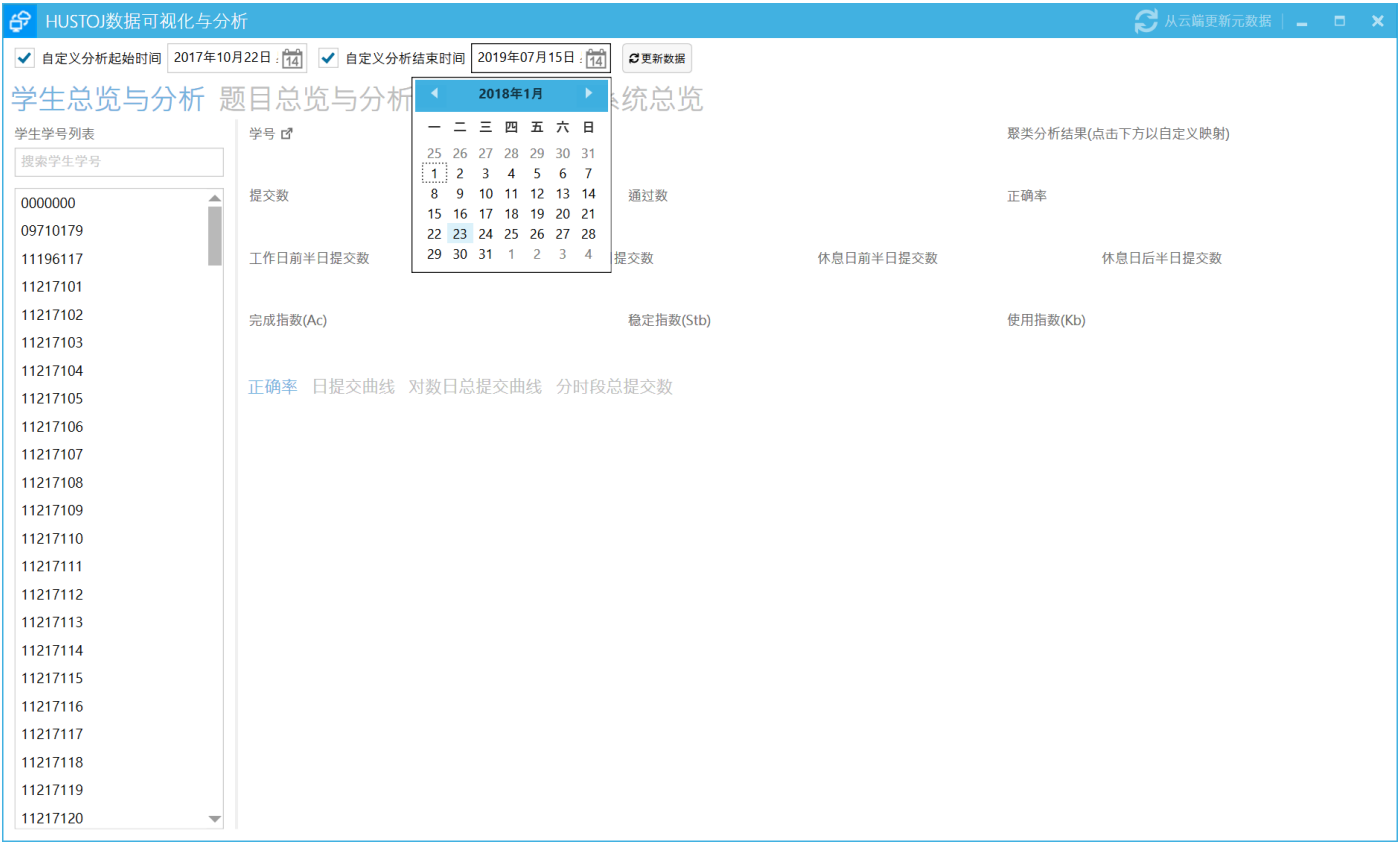


图 3.4 自定义分析结束时间

### 3.4 更新数据

您可以点击窗口右上角的 [从云端更新元数据] 按钮，从远程服务器更新日志数据并重新执行分析。该功能要求计算机拥有有效的网络连接。



图 3.5 更新数据时的界面

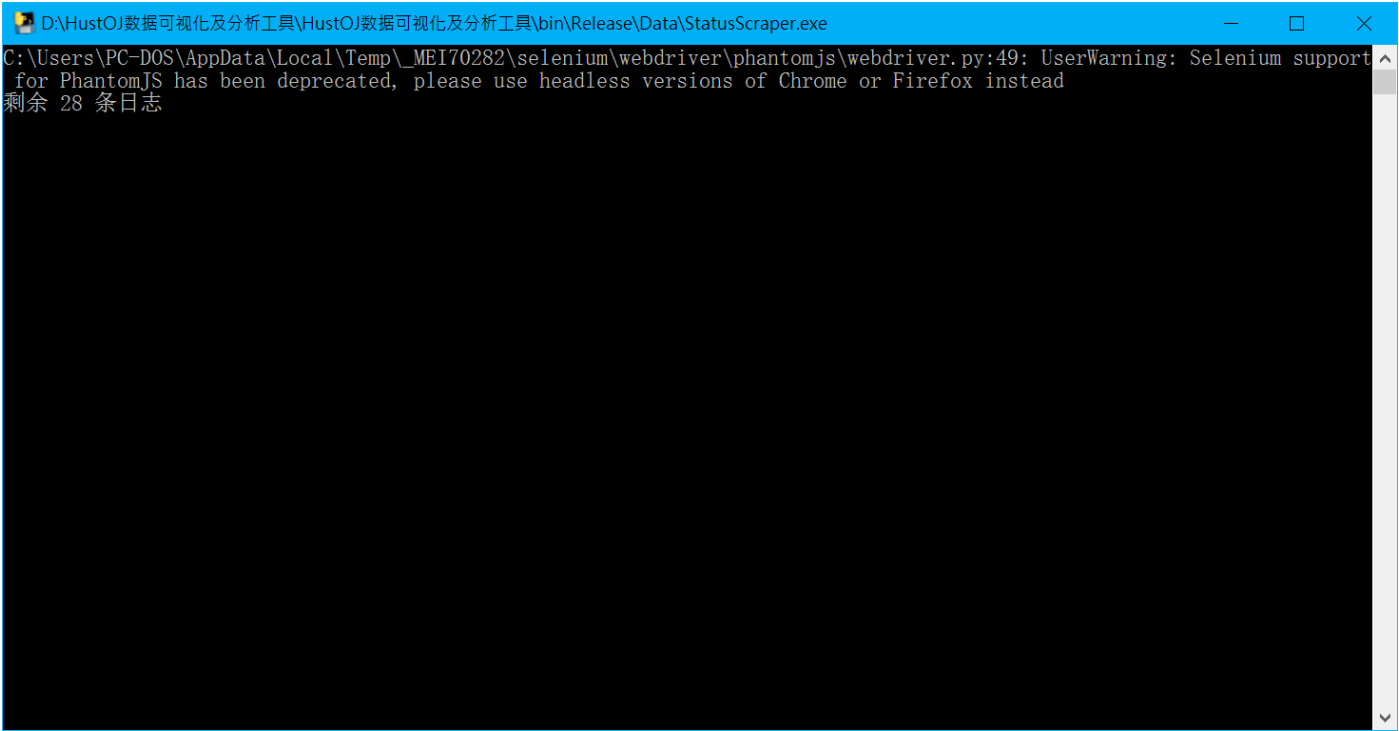


图 3.6 负责执行数据更新操作的外部程序

3.5 导出数据

您可以通过软件窗口右上角的 [导出分析数据] 按钮，导出原始数据或经过分析的数据。

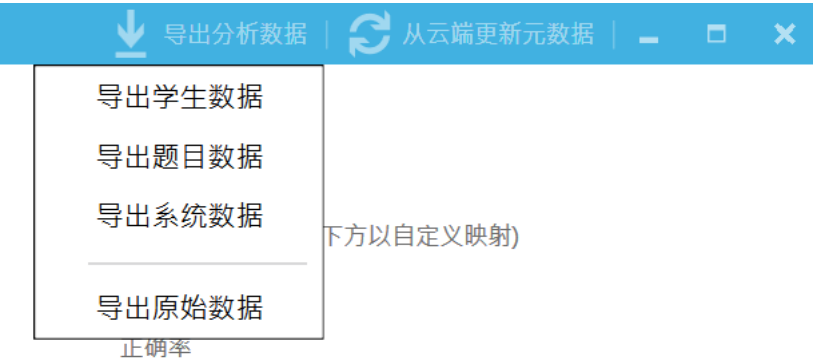


图 3.7 [导出分析数据] 菜单

3.5.1 导出原始数据

[导出原始数据] 选单允许您导出 CSV 格式的原始数据到指定的文件夹中。导出的文件及其用途如表 3.2 所示。

表 3.2 [导出原始数据] 产生的文件清单

文件名	Log.csv	User.csv	Problem.csv	Test.csv
文件内容	原始日志数据	用户清单	题目清单	测试清单

各数据文件内各列含义如表 3.3 至表 3.6 所示。

表 3.3 Log.csv 各列含义

列名	DATE_TIME	USER_ID	TEST_ID	PROBLEM_ID	RESULT
含义	日志产生的精确到秒的时间，格式为 YYYY-MM-DD HH:mm:ss	用户名	测试号	题号	是否通过该题 1 表示通过 0 表示未通过

表 3.4 User.csv 各列含义

列名	USER_ID
含义	用户名

表 3.5 Problem.csv 各列含义

列名	PROBLEM_ID	TEST_ID
含义	题号	该题所属测试号

表 3.6 Test.csv 各列含义

列名	TEST_ID	BEGIN_DATE	END_DATE
含义	测试序号，从 1 开始	测试首个提交出现时间，格式为 YYYY-MM-DD	测试最后一个提交出现时间，格式为 YYYY-MM-DD

### 3.5.2 导出学生数据

您可以通过 [导出学生数据] 功能，以 JS 对象简谱 (JavaScript Object Notation, JSON) 格式导出选定的经分析的学生数据。

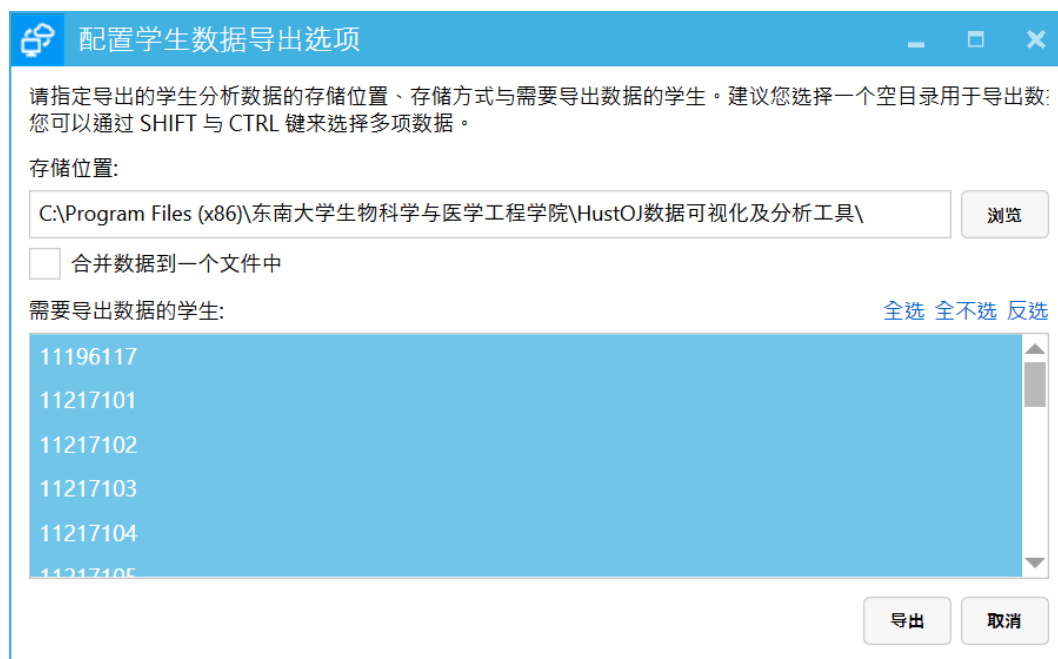


图 3.8 [导出学生数据] 配置界面

导出得到的 JSON 文件内各数据项名称、所属层级、数据类型及其含义如表 3.7 所示。

表 3.7 学生数据 JSON 文件各数据项、层级、数据类型及含义

数据项名	层级	数据类型	含义
AnalyzeStartDate	0	Date	用户定义的分析起始日期。
AnalyzeEndDate	0	Date	用户定义的分析结束日期。
IsDataMerged	0	Boolean	指示数据是否已合并，即指示 StudentDataSet 是否为 List 形式。
StudentDataSet	0	List(Of OJStudentInfo) 或 OJStudentInfo	学生数据（集），取决于获取的 JSON 文件是否是合并导出的。
StudentIDNumber	1	String	学生学号。
SubmitCount	1	Integer	总提交数。
ACCount	1	Integer	总通过数。
SubmitCountByDay	1	Dictionary(Of Date, Integer)	每日提交数，仅包含有提交的日期。

(续表)

数据项名	层级	数据类型	含义
SubmitCount-OnWorkdayAM	1	Integer	工作日上午提交总数。
SubmitCount-OnWorkdayPM	1	Integer	工作日下半日提交总数。
SubmitCount-OnRestdayAM	1	Integer	休息日上午提交总数。
SubmitCount-OnRestdayPM	1	Integer	休息日下半日提交总数。
SubmitCountByHour	1	Array(0 To 23)	每小时提交总数。
FittingAC	1	Double	完成指数 Ac。 $Ac = Pac + \alpha$ ，Pac 为通过数 ACCount， $\alpha$ 为通过率 ACCount/SubmitCount。
FittingR_Stb	1	Double	稳定指数 Stb。即拟合曲线的调整 R 平方值。
FittingK_Kb	1	Double	使用指数 Kb。即拟合曲线的斜率。
FittingB	1	Double	拟合曲线的截距。
ClustResult	1	Double	聚类分析结果。0 表示错误。

### 3.5.3 导出题目数据

您可以通过 [导出题目数据] 功能，以 JS 对象简谱 (JavaScript Object Notation, JSON) 格式导出选定的经分析的题目数据。

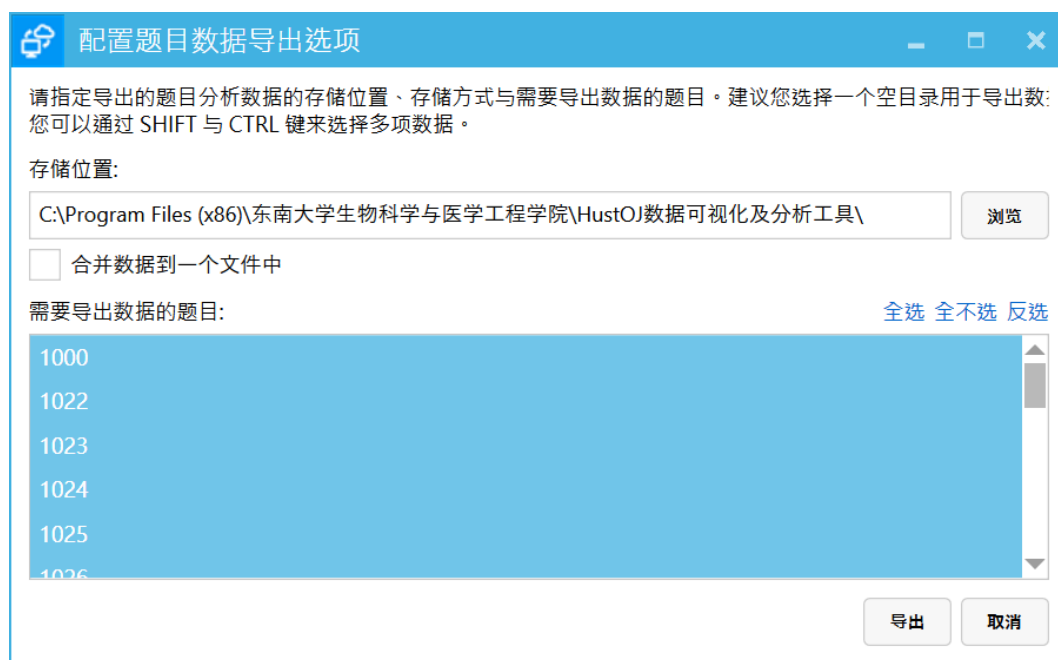


图 3.8 [导出题目数据] 配置界面

导出得到的 JSON 文件内各数据项名称、所属层级、数据类型及其含义如表 3.8 所示。



表 3.8 题目数据 JSON 文件各数据项、层级、数据类型及含义

数据项名	层级	数据类型	含义
AnalyzeStartDate	0	Date	用户定义的分析起始日期。
AnalyzeEndDate	0	Date	用户定义的分析结束日期。
IsDataMerged	0	Boolean	指示数据是否已合并，即指示 ProblemDataSet 是否为 List 形式。
ProblemDataSet	0	List(Of OJProblemInfo) 或 OJProblemInfo	题目数据（集），取决于获取的 JSON 文件是否是合并导出的。
ProblemIDNumber	1	String	题号。
ProblemTask- SequenceIndex	1	Integer	题目所属测试（题目集）号。
CreateTime	1	Date	题目创建时间，以第一个提交时间为准。
SubmitCount	1	Integer	总提交数。
ParticipantCount	1	Integer	总参与人数。
ACCount	1	Integer	总通过数。
EffortValue_Jq	1	Double	付出指数 Jq。
ParticipateValue_Eq	1	Double	参与指数 Eq。
SubmitCountByDay	1	Dictiondary(Of Date, Integer)	每日提交数，仅包含有提交的日期。
ClustResult	1	Integer	聚类分析结果，0 表示出错。

### 3.5.4 导出系统数据

您可以通过 [导出系统数据] 功能，以 JS 对象简谱 (JavaScript Object Notation, JSON) 格式导出经分析的系统数据。导出得到的 JSON 文件内各数据项名称、所属层级、数据类型及其含义如表 3.9 所示。

表 3.9 系统数据 JSON 文件各数据项、层级、数据类型及含义

数据项名	层级	数据类型	含义
LogStartDate	0	Date	日志起始日期。
LogEndDate	0	Date	日志结束日期。
AnalyzeStartDate	0	Date	用户定义的分析起始日期。
AnalyzeEndDate	0	Date	用户定义的分析结束日期。
IsDataMerged	0	Boolean	指示数据是否已合并，即指示 StudentDataSet 是否为 List 形式。
NewProblemCount	0	Dictiondary(Of Date, Integer)	每日新题目数，仅包含有新题目的日期。
SubmitCountByHour	0	Array(0 To 23)	每小时提交总数。

## 四、附录

### 4.1 软件原理

本软件原理请参考论文《基于在线评测系统的编程实战数据挖掘》。